DOCUMENT RESUME

ED 069 748 TM 002 203

AUTHOR	Lord, Frederic M.; Stocking, Martha
TITLE	Automated Hypothesis Tests and Standard Errors for Nonstandard Problems with Description of Computer Package: A Draft.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	ETS-RB-72-42
PUB DATE	Aug 72
NOTE	23p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Bulletins; *Computer Programs; *Electronic Data Processing; *Hypothesis Testing; *Mathematical Models; Research Methodology; *Standard Error of Measurement; Statistical Analysis; Testing; Tests of Significance

ABSTRACT

A general Computer program is described that will compute asymptotic standard errors and carry out significance tests for an endless variety of (standard and) nonstandard large-sample statistical problems, without requiring the statistician to derive asymptotic standard error formulas. The program assumes that the observations have a multinormal distribution and that the null hypothesis to be tested has the form xi = 0 where xi is some function (to be specified by the user) of means, variances, and covariances. Only minor programming is required to replace either or both of these assumptions. The package performs the automated hypothesis testing and consists of a main program and six subroutines. The package is written in Fortran IV. (Author/ON)

ED 069748

RESEARCH BULLETIN

TM 002 203

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RB-72-42

AUTOMATED HYPOTHESIS TESTS AND STANDARD ERRORS FOR NONSTANDARD PROBLEMS

Frederic M. Lord

WITH DESCRIPTION OF COMPUTER PACKAGE

Martha Stocking

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
August 1972

Automated Hypothesis Tests and Standard Errors for Nonstandard Problems

Frederic M. Lord

Abstract

A general computer program is described that will compute asymptotic standard errors and carry out significance tests for an endless variety of (standard and) nonstandard large-sample statistical problems, without requiring the statistician to derive asymptotic standard error formulas. The program assumes that the observations have a multinormal distribution and that the null hypothesis to be tested has the form $\xi = 0$ where ξ is some function (to be specified by the user) of means, variances, and covariances. Only minor reprogramming is required to replace either or both of these assumptions.

Automated Hypothesis Tests and Standard Errors for Nonstandard Problems*

Frederic M. Lord

Introduction

A general computer program can be written that will compute asymptotic standard errors and carry out significance tests for an endless variety of (standard and) nonstandard large-sample statistical problems, without requiring the statistician to derive asymptotic standard error formulas. This report describes one such computer program (currently available from the writer). The program, written by M. Stocking, is described starting on page 10.

As presently implemented, the program assumes that the observations have a multinormal distribution and that the null hypothesis to be tested has the form $\xi = 0$ where ξ is some function (to be specified by the user) of means, variances, and covariances. As will be pointed out, only minor reprogramming is required to replace either or both of these assumptions.

The present program is not set up to deal with vector hypotheses of the form $\underset{\sim}{\xi} = 0$. Possible program changes to accommodate vector hypotheses will be obvious to the statistician (see Moran, 1970, section 3).

Without user action, the program accommodates two samples, each composed of any number of observations on a maximum of 10 random variables. More samples (up to 20) with fewer random variables can be accommodated if the user sets all population covariances between variables from different samples equal to zero. In addition, the maximum of 10 random variables per sample can be increased, if desired.

---

*Research reported in this paper has been supported by grant GB-32781X from National Science Foundation.

Shenton, Bowman, and Sheehan (1971) describe computer procedures much more powerful and more versatile than those described here, with the exception that their program deals with univariate problems only.

Asymptotic Hypothesis Testing

Consider a sample consisting of a matrix $\underset{\sim}{X}$ of observations drawn from the distribution $f(\underset{\sim}{X}|\xi,\underset{\sim}{\theta})$, where $\underset{\sim}{\theta}$ is a vector of "nuisance" parameters mathematically independent of ξ and of each other. We wish to test the composite hypothesis H_0: $\xi = 0$. Let $\hat{\xi}$ and $\hat{\underset{\sim}{\theta}}$ be the maximum likelihood estimates of ξ and $\underset{\sim}{\theta}$ obtained without the restriction $\xi = 0$. H_0 can usually be tested by computing $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$ where $\hat{\sigma}_{\hat{\xi}}$ is the asymptotic sampling variance of $\hat{\xi}$ with $\hat{\xi}$ and $\hat{\underset{\sim}{\theta}}$ substituted for the unknown parameters ξ and $\underset{\sim}{\theta}$. The rejection region for H_0 consists of one or both tails of the asymptotic distribution of $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$ under H_0. Under regularity conditions (Cramér, 1946, section 33.3), this distribution is normal with zero mean and unit variance.

In a well-defined sense (Wald, 1943; Moran, 1970), the test just described is asymptotically optimal (locally asymptotically most powerful). If $\hat{\xi}$ and $\hat{\underset{\sim}{\theta}}$ are consistent estimators but not maximum likelihood estimators, the test will have the same significance level, but the power of the test (locally) may be low. A regularity condition worth noting is that $\xi = 0$ must not be a boundary of the range of ξ.

Input

In the present implementation two alternative methods are provided for data input. If the sample means $\underset{\sim}{m} \equiv \{m_i\}$ and the sample variance-

covariance matrix $\underset{\sim}{S} \equiv \|s_{ij}\|$ are available, these may be used as input. Otherwise, the raw observations are input and $\underset{\sim}{m}$ and $\underset{\sim}{S}$ automatically computed.

In order to use the packaged computer program, the statistician must specify what function $\xi = \xi(\underset{\sim}{\tau})$ or $\xi \equiv \xi(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ is to be tested, where ξ is differentiable and $\underset{\sim}{\tau}$ is a vector whose elements $\tau_g \equiv \tau_g(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ are differentiable functions of the mean vector $\underset{\sim}{\mu} \equiv \{\mu_i\}$ and the variance-covariance matrix $\underset{\sim}{\Sigma} \equiv \|\sigma_{ij}\|$ of the multinormal population. (In some problems--for example, in testing hypotheses about correlation coefficients--it is often convenient for the statistician to work with $\underset{\sim}{\tau}$, rather than with $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ only.) Note that $\hat{\xi} = \xi(\hat{\underset{\sim}{\tau}})$ or $\hat{\xi} = \xi(\hat{\underset{\sim}{\mu}},\hat{\underset{\sim}{\Sigma}})$ and that $\hat{\tau}_g = \tau_g(\hat{\underset{\sim}{\mu}},\hat{\underset{\sim}{\Sigma}})$ where each "hat" denotes a maximum likelihood estimate (for a rigorous treatment, see Zehna, 1966). In the present implementation, the statistician specifies the function ξ simply by writing a FORTRAN arithmetic assignment statement. On the left of the equal sign he writes XIHAT and on the right he writes down the explicit arithmetic expression for $\xi(\hat{\underset{\sim}{\tau}})$ or for $\xi(\hat{\underset{\sim}{\mu}},\hat{\underset{\sim}{\Sigma}})$. He inserts this FORTRAN statement at a pre-arranged point in a subroutine.

If $\hat{\underset{\sim}{\mu}} = \underset{\sim}{m}$ and $\hat{\underset{\sim}{\Sigma}} = \underset{\sim}{S}$, as in many simple problems, the user simply provides the explicit arithmetic expression for $\xi(\hat{\underset{\sim}{\mu}},\hat{\underset{\sim}{\Sigma}})$ and the computer proceeds directly to compute $\hat{\xi}$, $\hat{\sigma}_{\hat{\xi}}$, $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$, and finally the percentile at which $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$ falls in a standard normal distribution. If the user defines ξ in terms of the functions τ_g, then he must supply FORTRAN arithmetic assignment statements defining each $\hat{\tau}_g$ as a function of other $\hat{\tau}_g$ and the sufficient statistics $\underset{\sim}{m}$ and $\underset{\sim}{S}$.

If there are some restrictions on the parameters (other than that $\xi = 0$ under H_0), then $\hat{\underset{\sim}{\mu}} \neq \underset{\sim}{m}$ or $\hat{\underset{\sim}{\Sigma}} \neq \underset{\sim}{S}$. Since $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are estimated without the restriction $\xi = 0$, $\hat{\mu}$ and $\hat{\Sigma}$ will usually have a simpler mathematical form than would estimates obtained with this restriction. Whenever $\hat{\mu}_i \neq m_i$ or $\hat{\sigma}_{ij} \neq s_{ij}$, the statistician must insert in the subroutine arithmetic assignment statements defining $\hat{\mu}_i$ or $\hat{\sigma}_{ij}$ as functions of $\underset{\sim}{m}$ and $\underset{\sim}{S}$, or if convenient as functions of $\hat{\mu}$'s and $\hat{\sigma}$'s.

If formulas for efficient estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are not known to the statistician, he may substitute other consistent estimators, in which case (under regularity conditions) the significance test may be a conservative one in the sense that it rejects the null hypothesis less often than it should.

Automated Procedure

The program input described in the preceding section enables the computer to compute $\hat{\xi}$ for any values of $\underset{\sim}{m}$ and $\underset{\sim}{S}$. The main virtue of the program is that the statistician does not have to derive an explicit formula for the asymptotic sampling variance of $\hat{\xi}$. This is avoided because the computer uses the general formula

$$\hat{\sigma}^2_{\hat{\xi}} = \sum_i \sum_j \frac{\partial \hat{\xi}}{\partial m_i} \frac{\partial \hat{\xi}}{\partial m_j} \operatorname{cov}(m_i, m_j) + \sum_g \sum_h \sum_i \sum_j \frac{\partial \hat{\xi}}{\partial s_{gh}} \frac{\partial \hat{\xi}}{\partial s_{ij}} \operatorname{cov}(s_{gh}, s_{ij}) \quad , \tag{1}$$

where cov denotes a covariance with sample estimators substituted for population values. The required derivatives are approximated numerically

by the computer (see Description of Computer Package, p. 11), using only the formula for ξ, without need for symbolic differentiation (the user should check that ξ is differentiable, however). The necessary sampling covariances are automatically computed from standard formulas built into the computer program:

$$\left.\begin{aligned} \operatorname{cov}(m_i, m_j) &= \hat{\sigma}_{ij}/N \quad , \\ \operatorname{cov}(s_{gh}, s_{ij}) &= (\hat{\sigma}_{gi}\hat{\sigma}_{hj} + \hat{\sigma}_{gj}\hat{\sigma}_{hi})/N \quad , \end{aligned}\right\} \tag{2}$$

where N is the number of observations, and either $\hat{\sigma}_{ij} = s_{ij}$ or else an arithmetic assignment statement defining $\hat{\sigma}_{ij}$ has been provided in the program by the statistician.

If some parameters and statistics other than means, variances, and covariances are to be used and some distribution other than multinormal is to be assumed, it is only necessary to provide appropriate formulas for computing the maximum likelihood estimates and to insert appropriate formulas for sampling covariances in place of (2).

Illustrative Problems

The computer program has been checked out by applying it to numerical examples testing some two dozen different null hypotheses for which the numerical answers could be verified. A partial listing may suggest to the reader the scope of the program. Primes are used to distinguish parameters of two different populations; σ_i denotes the standard deviation of variable i.

Null Hypothesis	Restrictions
$\sigma_1 = \sigma_2$	$\sigma_{12} = 0$
$\rho_{12} = 0$	$\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$, $\rho_{12} = \sigma_{12}/\sigma_1\sigma_2$
$\frac{1}{2}[\log(1 + \rho) - \log(1 - \rho)] = \text{constant}$	$\rho = \sigma_{12}/\sigma_1\sigma_2$
$\rho_{12}\rho_{34} - \rho_{13}\rho_{24} = 0$	$\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$ $(i,j = 1, 2, 3, 4)$
$\sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24} = 0$	$\sigma_{13} = \sigma_{14}$, $\sigma_{23} = \sigma_{24}$, $\sigma_3 = \sigma_4$
$\frac{\rho_{12}\rho_{34} - \rho_{13}\rho_{24}}{\sqrt{(\rho_{34} - \rho_{13}^2)}\sqrt{(\rho_{34} - \rho_{24}^2)}} = 0$	same
$\mu_1' = 0$ (both populations are bivariate)	$\mu_2' = \mu_2''$, $\sigma_2' = \sigma_2''$ (Lord, 1955)
trace Σ^{-1} = constant	none
$\lvert\Sigma\rvert$ = constant	none (Madansky and Olkin, 1969)
trace$(\Sigma'^{-1}\Sigma'')$ = constant	none

Example

Consider testing the null hypothesis that the tetrad $\rho_{12}\rho_{34} - \rho_{13}\rho_{24} = 0$ where $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$, with the restriction that $\sigma_3 = \sigma_4$, $\sigma_{13} = \sigma_{14}$, $\sigma_{23} = \sigma_{24}$ (Lord, 1972). First, let us replace the null hypothesis by the equivalent hypothesis that $\xi \equiv \sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24} = 0$. This definition of the function ξ is provided to the computer by inserting in the program the arithmetic assignment statement

XIHAT = [FORTRAN equivalent of $\hat{\sigma}_{12}\hat{\sigma}_{34} - \hat{\sigma}_{13}\hat{\sigma}_{24}$] .

It is also necessary to provide FORTRAN statements defining the maximum likelihood estimators

$$\hat{\sigma}_{13} = \hat{\sigma}_{14} = (s_{13} + s_{14})/2 \quad ,$$

$$\hat{\sigma}_{23} = \hat{\sigma}_{24} = (s_{23} + s_{24})/2 \quad ,$$

$$\hat{\sigma}_3^2 = \hat{\sigma}_4^2 = (s_3^2 + s_4^2)/2 \quad .$$

The estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_{12}$ are not to be defined explicitly in the program, with the result that the computer will assume (correctly) that $\hat{\sigma}_1^2 = s_{11}$, $\hat{\sigma}_2^2 = s_{22}$, $\hat{\sigma}_{12} = s_{12}$, and $\hat{\sigma}_{34} = s_{34}$.

Provided with the four FORTRAN definitions shown, the computer will now compute $\hat{\xi}$, its estimated asymptotic variance $\hat{\sigma}_{\hat{\xi}}$, the test statistic $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$, and the percentile corresponding to the test statistic. All this is easier for the statistician than deriving the formula for $\hat{\sigma}_{\hat{\xi}}^2$ --an eighth degree multinomial containing ten terms--and then computing the test statistic from this formula.

In a Monte Carlo study, 1000 values of $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$ and their probability levels were computed using the program described, where

$$\hat{\xi} = \frac{\hat{\rho}_{12}\hat{\rho}_{34} - \hat{\rho}_{13}\hat{\rho}_{24}}{\sqrt{(\hat{\rho}_{34} - \hat{\rho}_{13}^2)}\sqrt{(\hat{\rho}_{34} - \hat{\rho}_{24}^2)}} \quad .$$

The time required on a 360/65 for all 1000 was about 80 seconds.

References

Cramér, H. (1946), Mathematical methods of statistics. Princeton, N.J.: Princeton University Press.

Lord, F. M. (1955), Estimation of parameters from incomplete data, J. Amer. Statist. Assoc. 50, 870-876.

Lord, F. M. (1972), Significance test for a partial correlation corrected for attenuation. Research Bulletin 72-6 and ONR Technical Report Contract No. N00014-69-C-0017. Princeton, N.J.: Educational Testing Service.

Madansky, A. and Olkin, I. (1969), Approximate confidence regions for constraint parameters. In P. R. Krishnaiah (Ed.), Multivariate analysis--II. New York: Academic Press.

Moran, P. A. P. (1970), On asymptotically optimal tests of composite hypotheses, Biometrika 57, 47-55.

Shenton, L. R., Bowman, K. O., and Sheehan, D. (1971), Sampling of moments associated with univariate distributions, J. Roy. Statist. Soc., Ser. B, 33, 444-457.

Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Amer. Math. Soc. 54, 426-482.

Zehna, P. W. (1966), Invariance of maximum likelihood estimators, Ann. Math. Statist. 37, 744.

Description of Computer Package

Martha Stocking

1. General Description

The package that performs the automated hypothesis testing consists of a main program and six subroutines. The package is written in FORTRAN IV in double precision and when compiled on an IBM 360/65 using the G compiler, takes 56K. This size will be increased if the user either enlarges the number of variables in a sample (see section 5) or modifies subroutine XIHAT substantially (see section 4).

LASAHT	Main program for Large Sample Hypothesis Testing. LASAHT reads input, computes derivatives, and calls subroutines COVM, COVS, IN, MATOUT, and XIHAT.
IN	Subroutine called by LASAHT if input is raw data to read data and compute means and covariances.
COVM	Subroutine called by LASAHT to compute first sum in equation (1).
COVS	Subroutine called by LASAHT to compute second sum in equation (1).
MATOUT	Subroutine called by LASAHT to print upper triangular matrices.
FILL	Subroutine called by MATOUT.
XIHAT	Subroutine in which user will supply formulas for $\hat{\xi}$ and any restrictions (see section 4).

2. Computation of Partial Derivatives

The partial derivatives are computed numerically. All elements in $\hat{\underset{\sim}{\tau}}$ are held constant except one. This element is increased by some increment, and a value for $\hat{\xi}$ is computed; the element is then decreased from its original value by the same increment, and another value of $\hat{\xi}$ is computed. The derivative is approximated as the difference between these two values of $\hat{\xi}$ divided by twice the increment.

In computations with fixed word length, the optimum size of the increment is not known. A repetitive procedure is used in which the size of the increment is decreased up to the point where the difference between two successively computed approximations to the derivative increases (due to truncation error). The first value used for the increment is 0.001. In successive repetitions, this is decreased by a factor of 10, up to a maximum of five times. If the differences between successive approximations do not increase, the last derivative computed is used.

3. Input, Output, Program Messages for User

Input

The description of the input for one or two samples with a maximum of 10 variables each is simple, and will be discussed first. The description for more samples with fewer variables is more complex and will be discussed later. The following is a description of the input cards required. The user must follow the normal FORTRAN convention of right-adjusting all values within the specified fields.

CARD 1 FORMAT(16I5)

cols. 1-5	NSAMP	The number of independent samples of data. Must be one or two.

CARD 2 FORMAT(10A8)

cols. 1-80	TITLE	Title of first sample

CARD 3 FORMAT(16I5)

cols. 1-5	NOBS1	Number of observations in first sample
cols. 6-10	N1	Number of variables in first sample, maximum of 10
cols. 11-15	INPUT	= 0 , input for first sample is raw data = 1 , input for first sample is sample means and sample variance-covariance matrix.

CARD 4 FORMAT(10A8)

cols. 1-80	FMT	Variable format for reading data for first sample. Will be used for either raw data or means and variance-covariance matrix, depending upon cols. 11-15 of card 3. Must specify a floating-point format.

If INPUT = 0

CARD 5+ FORMAT(FMT)

cols. 1-80	RAW(I),(I=1,...,N1)	Observations on N1 variables, one record per set of N1 observations, with each set of observations beginning a new record. Punched according to format specified on card 4.

If INPUT = 1

CARD 5+ FORMAT(FMT)

cols. 1-80 X(I),(I=1,...,N1) N1 sample means for first set of data, punched according to format specified on card 4.

CARD 6+ FORMAT(FMT)

cols. 1-80 S(I,J),(I=1,...,N1; J=I,...,N1) Upper triangle, including the diagonal, of the sample variance-covariance matrix. This triangular matrix must be punched row by row (each row begins with the diagonal element) with each row beginning a new record according to format specified on card 4.

If cols. 1-5 of card 1 = 2, i.e., there are two samples, then cards 2 through 6+ are repeated for the second sample. The variable names which change, and their meanings are:

NOBS2 - number of observations in the second sample

N2 - number of variables in the second sample, maximum of 10

X2 - means for second sample

S2 - variance-covariance matrix for second sample.

All other variables are reused for the second sample.

It is possible to increase the number of samples up to a maximum of 20, but one must also reduce the number of variables per sample, and specify that the covariances between variables from different samples are zero. The input cards remain essentially the same. In particular, card 1, NSAMP, must still be either one or two.

Suppose we have two samples, each with one variable, i.e., two univariate samples. If NOBS1 = NOBS2 , exactly the same results can be obtained from the program by considering the data as one bivariate sample with $s_{12} = 0$. The input cards for two univariate samples with sample means and covariances as input would be similar to

Card 1	cols. 1-5	NSAMP = 2
Card 2	cols. 1-80	TITLE
Card 3	cols. 1-5	NOBS1
	cols. 6-10	N1 = 1
	cols. 11-15	INPUT = 1
Card 4	cols. 1-80	FMT = (8F10.4) (say)
Card 5	cols. 1-10	X(1)
Card 6	cols. 1-10	S(1,1)
Card 2*	cols. 1-5	TITLE for 2nd sample
Card 3*	cols. 1-5	NOBS2
	cols. 6-10	N2 = 1
	cols. 11-15	INPUT = 1
Card 4*	cols. 1-80	FMT = (8F10.4)
Card 5*	cols. 1-80	X2(1)
Card 6*	cols. 1-10	S2(1,1)

*Indicates second sample.

Considering these data as one bivariate sample we would have

Card 1	cols. 1-5	NSAMP = 1
Card 2	cols. 1-80	TITLE
Card 3	cols. 1-5	NOBS1
	cols. 6-10	N1 = 2

Card 4	cols. 1-80	FMT = (8F10.4)
Card 5	cols. 1-10	X(1)
	cols. 11-20	X(2)
Card 6(a)	cols. 1-10	S(1,1)
	cols. 11-20	S(1,2) = 0
Card 6(b)	cols. 1-10	S(2,2)

Further examples might be of some use here. Suppose the user has four samples of the same size, each with two variables. This can be handled by the program by considering these data to be one sample (NSAMP = 1) with eight variables and the following sample variance-covariance matrix

$$\begin{matrix} s_{11} & s_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ & s_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ & & s_{33} & s_{34} & 0 & 0 & 0 & 0 \\ & & & s_{44} & 0 & 0 & 0 & 0 \\ & & & & s_{55} & s_{56} & 0 & 0 \\ & & & & & s_{66} & 0 & 0 \\ & & & & & & s_{77} & s_{78} \\ & & & & & & & 88 \end{matrix} .$$

Or we can consider these data as two samples (NSAMP = 2) with 4 variables each and obtain the same result. The sample variance-covariance matrix for the first sample would be

$$\begin{matrix} s_{11} & s_{12} & 0 & 0 \\ & s_{22} & 0 & 0 \\ & & s_{33} & s_{34} \\ & & & s_{44} \end{matrix} ,$$

and for the second sample

$$\begin{matrix} s^*_{11} & s^*_{12} & 0 & 0 \\ & s^*_{22} & 0 & 0 \\ & & s^*_{33} & s^*_{34} \\ & & & s^*_{44} \end{matrix}$$

* = second sample

By extension of these examples, it is clear that without changing the program, a maximum of 20 univariate samples can be handled if they are all of the same size. In this case, we would treat this as two (NSAMP = 2) samples, each with 10 variables. The sample variance-covariance matrix would be similar for each sample--the main diagonal would contain the variances, and all off-diagonal elements would be zero.

Output

The output from the program is largely self-explanatory. A title is printed at the top of the first page. The number of samples (NSAMP) is also printed, and will be either one or two. It will not reflect the actual number of samples if the user has taken advantage of the facilities described in the input section to expand the number of samples. Also, the time and date of the run are printed (see section 6).

For each of two possible samples there is printed the title, number of observations, number of variables, and the variable INPUT. In addition, to provide the user with means of checking his input, the sample means, and upper right triangle of the sample variance-covariance matrix is printed.

The estimated means and variance-covariance matrix are then printed. These values will be identical to the sample values except for those values for which the user has provided formulas for computing different estimates in subroutine XIHAT (see section 4). This output is also provided for user verification of his formulas.

The value of $\hat{\xi}$, computed with the estimated values, is then printed. In addition, the partial derivatives of $\hat{\xi}$ with respect to the means and variance-covariance matrices are printed for each sample.

Finally, $\hat{\xi}$ is printed again, along with $\text{var}(\hat{\xi})$, $\hat{\sigma}_{\hat{\xi}} = \sqrt{\text{var}(\hat{\xi})}$, $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$, and the percentile at which $\hat{\xi}/\hat{\sigma}_{\hat{\xi}}$ falls in a standard normal distribution.

Messages

The program detects two conditions that will result in the printing of a message:

1. If NSAMP is greater than two, the actual value of NSAMP that the user supplied is printed and the program exits.

2. The program uses the value of the last derivative computed as the partial derivative, and continues on to the computation of the next partial derivative. If a derivative is obtained from computation with the smallest increment, a message is printed.

4. Subroutine XIHAT

Subroutine XIHAT is the subroutine in which the user inserts coding for his specific null hypothesis $\hat{\xi} = 0$. If the user wishes to specify estimates other than sample means and variance-covariances, he also inserts the coding in this subroutine.

Variables containing information transmitted to this subroutine as arguments are as follows (these names differ slightly from the corresponding variable names in LASAHT):

Variable name	Dimension	Contents
X	10	Sample means for first sample
S	10 by 10	Sample variance-covariance matrix for first sample
X2	10	Sample means for second sample
S2	10 by 10	Sample variance-covariance matrix for second sample
N		Number of variables in first sample
N2		Number of variables in second sample
NOBS		Number of observations for first sample
NOBS2		Number of observations for second sample
NSAMP		Either 1 or 2, same variable as card 1 of input

Variables containing information transmitted from this subroutine as arguments are as follows:

M	10	Estimated means for first sample
$	10 by 10	Estimated variance-covariance matrix for first sample
M2	10	Estimated means for second sample
$2	10 by 10	Estimated variance-covariance matrix for second sample
XIHAT		The value of $\hat{\xi}$ computed in this subroutine

The FORTRAN statements down to 9999 transfer the sample means and covariances into the estimated means and covariances. This insures that the estimated means and covariances will be the same as the sample values unless modified by the user.

If the user wishes to specify other estimates for some or all of the means and covariances, he must insert the FORTRAN code to do that operation between the first two asterisk cards. He must follow the following rules:

1. In referring to any element of any covariance matrix, the second subscript must be greater than or equal to the first, i.e., only the upper right triangle is to be worked with.
2. The estimates supplied must appear to the left of an equal sign.
3. The first line of FORTRAN code must not be a format statement, and must have the statement number 100.
4. Statement numbers 9998 and 9999 must not be used.

The user must insert the FORTRAN code for the computation of $\hat{\xi}$ between the second set of asterisk cards. This code may consist of a number of statements. The user may refer to any variable previously defined. The following rules must be followed:

1. In referring to any element of any variance-covariance matrix, the second subscript must be greater than the first.
2. Statement numbers 9998 and 9999 must not be used, and statement numbers used in specifying estimates (if any) must not be repeated.

3. If no estimates were specified, the first line of code must not be a format statement, and must have the statement number 100.

4. The variable XIHAT must appear to the left of an equal sign.

5. Instructions for Program Changes

Increasing the Number of Variables

The following vectors in the main program LASAHT have dimensions which are dependent upon the number of variables in each sample. Let n_1 indicate the maximum number of variables in the first sample, and n_2 be the maximum number in the second sample.

MSAVE(n_1), \$SAVE(n_1,n_1), M(n_1), \$(n_1,n_1), S(n_1,n_1), X(n_1),
M2SAVE(n_2), \$2SAVE(n_2,n_2), M2(n_2), \$2(n_2,n_2), S2(n_2,n_2), X2(n_2)

$$\text{DER}\left(\max\left(\frac{n_1(n_1+1)}{2}, \frac{n_2(n_2+1)}{2}\right)\right)$$

$$\text{TAU}(n_1 + n_2 + n_1^2 + n_2^2)$$

DOUT(max(n_1,n_2), max(n_1,n_2))

INDEX(max(n_1,n_2))

The equivalencing of TAU to MSAVE, M2SAVE, \$SAVE, and \$2SAVE is arranged so that MSAVE(1) is TAU(1), M2SAVE(1) is TAU(n_1 + 1), \$SAVE(1,1) is TAU(n_1 + n_2 + 1) and \$2SAVE(1,1) is TAU(n_1 + n_2 + n_1^2 + 1).

The user may change the maximum number of variables by using the above formulas for determining the correct dimension and equivalence statements. There are only three restrictions: (1) the maximum number of variables must be the same for both samples; (2) the variable IDEM must be set

equal to the maximum number of variables (see ISN 14 in the listing); (3) all subroutine dimension statements must be changed to reflect the new dimensions in LASAHT.

Changing Derivative Computation

The algorithm to compute the partial derivatives required is partially described in section 2. The actual details of the algorithm will not be discussed here. If the user wishes to change the initial value of the increment, he supplies a different value for DEL(ISN 105). If the user wishes to change the number of repetitions, he changes ISN 132 to test IDEL (iteration counter) against the maximum number of iterations.

6. Machine Dependent Features

There are a number of features used in this program which are peculiar to FORTRAN on an IBM 360 in addition to the 4 byte word, hexadecimal machine. Almost all variables which are normally floating-point are implicitly defined as double-precision floating point numbers. Three machine subroutines are called: (1) ERRSET to suppress the printing of end of file messages, (2) SETCLK to set the interval timer for this program, (3) DATET to get date and time of day for printing an output.

Any or all of these features may be eliminated by the user (making appropriate modifications of statement numbers, etc.) or replaced.